Published in final edited form as:

Psychol Med. 2016 July ; 46(10): 2059–2069. doi:10.1017/S0033291716000453.

Genome-wide SNP Heritability of Nicotine Dependence as a Multidimensional Phenotype

LC Bidwell^{1,2,3,*}, RHC Palmer^{2,3,*}, L Brick³, JE McGeary^{4,2,3}, and VS Knopik^{2,3}

¹Institute of Cognitive Science, University of Colorado at Boulder, Boulder, CO

²Department of Psychiatry and Human Behavior, Alpert Medical School of Brown University, Providence, RI

³Division of Behavioral Genetics, Department of Psychiatry, Rhode Island Hospital, Providence RI

⁴Department of Veterans Affairs, Providence VA Medical Center, Providence, RI

⁵Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder CO

Abstract

Background—Heritability estimates from twin studies of the multi-faceted phenotype of nicotine dependence (ND) range from moderate to high (31–60%), but vary substantially based on the specific ND-related construct examined. The current study estimated the aggregate role of common genetic variants on key ND constructs.

Methods—Genomic Relatedness Restricted Maximum Likelihood (GREML) was used to decompose phenotypic variance across multiple ND indices using 796,125 polymorphisms from 2346 unrelated “lifetime ever smokers” of European ancestry. Measures included DSM-IV ND and FTND summary measures and constituent constructs (e.g. withdrawal severity, tolerance, heaviness of smoking, and time spent smoking). Exploratory (EFA) and confirmatory (CFA) factor models were used to describe the covariance structure across ND measures; resulting factor(s) were the subject(s) of GREML analyses.

Results—Factor models indicated highly correlated DSM-IV and FTND factors for ND (0.545 [95% CI: 0.50–0.60]) that could be represented as a higher-order factor (NIC DEP). Additive genetic influences on NIC DEP was 33% (S.E. = .14, $p = .009$). Post hoc analyses indicated moderate genetic effects on the DSM-IV (34% [S.E. = .14, $p = .008$]) and FTND factors (26% [S.E. = .14, $p = .032$]), both of which were influenced by the same genetic effects, ($r_{G-SNP} = 1.00$, S.E. = .09, $p < .00001$).

Conclusions—Overall, common SNPs accounted for a large proportion of the genetic influences on ND-related phenotypes that have been observed in twin studies. Genetic contributions across distinct ND scales were largely influenced by shared genetic factors.

Corresponding Author: L. Cinnamon Bidwell, UCB 344, Boulder CO, 80309.

*Drs. Bidwell and Palmer contributed equally to this manuscript.

Conflict of Interest: None.

Keywords

Addiction; Substance Dependence; Genetics; SNP-heritability

Introduction

Nicotine dependence (ND) is a complex, multi-dimensional phenotype influenced by both genetic and environmental factors. Although the role of genetics in ND has been well established in twin and family studies, heritability estimates from these studies vary substantially based on the particular ND-related phenotype measured. Heritability estimates across studies examining nicotine-dependence phenotypes, as typically defined by the Diagnostic and Statistical Manual of Mental Disorders (DSM) – Third Edition, Revised (DSM-III-R; American Psychiatric Association, 1987) and DSM-IV (American Psychiatric Association, 1994) or Fagerström Tolerance Questionnaire (FTND; Heatherton *et al.*, 1989) criteria, range from 31% to 60% (Lesso *et al.*, 2004, McGue *et al.*, 2000, Pergadia *et al.*, 2006b, True *et al.*, 1999). More discrete ND measures, such as quantity of cigarettes smoked and heaviness of smoking, have been shown to be significantly heritable, with estimates of heritability ranging from 45% to 86% for number of cigarettes per day (CPD; Broms *et al.*, 2006, Hettema *et al.*, 1999, Kaprio *et al.*, 1981, Koopmans *et al.*, 1999, Lessov *et al.*, 2004, Swan *et al.*, 1990), 46% to 49% for heavy smoking (Pergadia *et al.*, 2006a, Swan *et al.*, 1997), and 59–71% for the Heaviness of Smoking Index (HSI; Heatherton *et al.*, 1989), which comprises two of the seven FTND items – time to first cigarette in the morning and number of CPD (Lesso *et al.*, 2004, Pergadia *et al.*, 2006b).

Gold-standard measures of ND are known to be multidimensional in their psychometric properties. For example, scales such as the FTND and scales comprising DSM-IV items actually have modest internal consistencies (scores on the items are not highly correlated with one another) (Piper *et al.*, 2006). There is also strong evidence that the most commonly administered dependence scales (e.g., the FTND and the DSM-IV criteria) are not highly correlated with each other (Breslau and Johnson, 2000, Moolchan *et al.*, 2002, Piper *et al.*, 2006). Moolchan and colleagues (2002) found that the kappa estimate of concordance was only 0.2, not much better than that expected by chance alone. This finding has been replicated (Kandel *et al.*, 2005) and it is largely accepted that the two measures, although both claiming to be assessing nicotine dependence, are, in fact, assessing two different aspects of ND and potentially two subsets of smokers. The DSM-based criteria appear to be broader, assessing the physiological symptoms of ND (tolerance and withdrawal), as well as a larger behavioral pattern typical of dependent use (e.g. continued use despite problems) and psychological aspects of dependence (e.g. impaired control over use; (Hughes, 2003). The FTND appears to have a narrower focus that emphasizes the behavioral components of dependence and some physiological components related to tolerance and withdrawal (e.g. time to first cigarette) and is more closely related to behavioral (Moolchan *et al.*, 2002) and biological (Payne *et al.*, 1994) indices of smoking, but less to psychological symptoms of dependence (Dijkstra and Tromp, 2002). Another consideration in ND assessment is that, although dependence measures often show lower internal consistency and minimal relations with one another (Hughes *et al.*, 2004, Piper *et al.*, 2006), factor analyses show that such

measures often load highly onto common factors. Even when factor analyses suggest multiple factors, the factors are highly inter-correlated (Lessov *et al.*, 2004, Piper *et al.*, 2004, Shiffman *et al.*, 2004). Taken together, research suggests that ND symptoms and features are best accounted for by a multidimensional model that includes multiple, correlated ND features such as withdrawal severity, smoking heaviness, time spent smoking, tolerance, and potentially other aspects (Baker *et al.*, 2009, Hudmon *et al.*, 2003, National Cancer Institute, 2009, Pomerleau *et al.*, 1993).

The heterogeneity and lack of agreement of gold-standard dependence measures, such as the DSM and the FTND, can be viewed as a useful opportunity for genetic research. Accepting the hypothesis that ND is multi-dimensional allows us to parse and distill the ND phenotype from a genetics perspective and to examine which aspects of ND have the highest genetic loading. For example, a study by Lessov and colleagues (2004) examined the relative contribution of genetic and environmental influences to the variance of individual ND criteria as well as a ND diagnosis defined by the DSM-IV and the HSI. The results showed substantial heritability across genders for DSM-IV nicotine tolerance (73%), withdrawal (53%), smoking more than intended (62%), and both HSI items—time to first cigarette in the morning (68%) and number of CPD (70%); estimates for ‘difficulty in quitting smoking’ varied across genders (68% in females, 54% in males). Relatively moderate heritability was been observed for DSM-IV items: spending a great deal of time smoking (45%), smoking despite physical or psychological problems (39%), and giving up important activities to smoke (26%). These results suggest that the DSM-IV criteria of tolerance, withdrawal, and difficulty in quitting smoking, and the HSI may be the most genetically influenced indicators of nicotine dependence in adults.

Despite these substantial genetic effects derived across aspects of ND using traditional twin approaches, molecular genetic and genome-wide association studies (GWAS) have had limited success in isolating or aggregating specific causal variants and accounting for the genetic variance estimates derived from twin studies. Although several GWAS studies have utilized novel phenotypic definitions, such as expired carbon monoxide and smoking frequency, in order to maximize the chances of finding genetic markers for ND (Bloom *et al.*, 2014, Loukola *et al.*, 2014), effect sizes still typically fall between 1–3%. In part to address the possibility that most variants truly associated with complex traits have effect sizes too small to detect individually using GWAS studies, Yang *et al.* (2013) developed Genomic-related-matrix Restricted Maxim Likelihood (GREML; implemented in the Genome-wide Complex Trait Analysis (GCTA) software) that focuses on the estimation of the phenotypic variance explained by genome-wide similarity at genotyped SNPs. Rather than testing each SNP individually, GREML decomposes the phenotypic variance into two components: (1) effects due to the additive influences of all measured autosomal SNPs (SNP heritability or h^2_{SNP}) and (2) the effects due to unmeasured environmental influences, random noise or the effects of genetic variants that were not measured by the genotyping array. This approach allows for an estimate of phenotypic variability explained by genome-wide SNP data. GREML has been used to estimate the additive genetic influences on DSM-III ND in one prior study, which reported an aggregate SNP heritability of 36% and suggested that common SNPs account for a relatively large portion of genetic effects detected using traditional biometrical twin modeling approaches (Vrieze *et al.*, 2013). Given

the multidimensional nature of ND, a critical next step is to examine the aggregate role of common SNPs on various aspects of ND, including both discrete and multidimensional ND phenotypes.

Current study

The current study aimed to identify genetic effects on multiple facets of ND, including broad ND constructs derived from gold-standard measures of ND and item-level assessments that tap more discrete, albeit core, features of ND, such as withdrawal severity, tolerance, heaviness of smoking, and time spent smoking. We hypothesized that common genetic variants account for at least half of the genetic variance observed in twin/adoption studies. Further, despite phenotypic heterogeneity, we hypothesized that a common set of genetic factors would account for the genetic variances identified across the definitions. To test these hypotheses, we first used factor analysis to replicate prior work and distill our multi-faceted ND phenotype. Secondly, we examined genome-wide additive genetic influences on ND at the discrete item level as well as on the observed ND factors. These measures were analyzed separately to explore how the magnitude of identified additive genetic effects differed across facets of ND. Finally, we used a bivariate model to test whether these gold-standard measures of ND index the same genetic liability.

Materials and Methods

Study population

The current study utilizes individuals of European ancestry (EA) who reported having ever used cigarettes (N=2346; average age = 38.41 (SD = 9.58), %male=45%) and were part of the Study of Addiction: Genetics and Environment (SAGE). Details on the SAGE cohort are available elsewhere (National Center for Biotechnology Information [NCBI], n.d.). SAGE is a collection of individuals from several case-control genetic studies and was designed for the purpose of identifying genetic factors that contribute to addiction. The current study included EAs with a history of smoking tobacco-based products and who also had information on DSM-IV nicotine dependence and the FTND. Participants in SAGE originate from three studies designed to study drug addiction (the Collaborative Study on the Genetics of Alcoholism, the Family Study on Cocaine Dependence, and the Collaborative Study on Nicotine Dependence). Although participants were ascertained to study alcohol, cocaine, and nicotine, cases were not excluded if they endorsed criteria for dependence on other substances (e.g., cannabis, opiates, stimulants, hallucinogens); controls were required to be alcohol, cocaine, or tobacco users without a history of drug dependence. Supplemental Table 1 summarizes the history of substance problems in the sample as well as the levels of comorbidity of substance problems in SAGE.

Measurements

Dependent measures were DSM-IV nicotine dependence and FTND symptoms gathered using the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (Bucholz *et al.*, 1994) or modified versions used within the sub-samples that comprise SAGE (i.e., the Semi-Structured Assessment of Nicotine Dependence used in the Collaborative Genetic Study of Nicotine Dependence (COGEN) and the Semi-Structured Assessment for

Cocaine Dependence used in the family Study of Cocaine Dependence (FSCD). Participants were included if they had a history of ever using tobacco daily for a month or more, including smoking cigarettes, cigars, or pipes, as well as using snuff/chewed tobacco. Of these individuals, 1331 met DSM-IV criteria for ND and 714 met the clinical cut FTND cut off of >4; 670 individuals met ND criteria on both DSM-IV and FTND measures, consistent with prior work suggesting DSM-IV and FTND emphasize different aspects of the ND phenotype and assign diagnoses to two different subsets of smokers.

In order to assess the common and specific aspects of DSM-IV and FTND measures of ND we utilized factor analysis of the individual criteria for each measure. In doing so, we sought to construct a composite score that reflected the shared variance across DSM-IV and FTND symptoms. This was achieved by using a factor analytic approach that included both a model-building phase and a data extraction phase. In the model-building phase, we first used exploratory factor analysis (EFA) on a random subset of all smokers (N=1180) using weighted least squares mean variance (WLSMV) estimation to adjust for dichotomization of DSM-IV items. In addition to retaining eigenvalues greater than one, we used Scree Plots and interpretability of loadings to identify the number of factors to retain. Second, we fitted two confirmatory factor analysis (CFA) models using the remaining half of the smokers (N=1166). The first model (CFA-Model-I) was based on the observed number of factors and factor loadings from the EFA. The second model (CFA-Model-II) tested the hypothesis that there exists two separate, but correlated, underlying latent traits separately indicated by DSM-IV and FTND. The fit of the CFA models was evaluated using standard EFA/CFA fit indices (i.e., root mean square error of approximation (RMSEA; value <0.08), comparable fit index (CFI; value >0.90), and Tucker-Lewis index (TLI; >0.90) (MacCallum *et al.*, 1996). We compared nested models by examining model fit indices and determining significant change in CFI ($CFI_{\text{constrained model}} - CFI_{\text{unconstrained model}} < -0.01$) (Cheung and Rensvold, 2002). In the data extraction phase, we obtained factor scores from the best fitting CFA model(s) from the model-building phase using the entire sample (N=2346). Factor scores from the data extraction phase were used in the genetic models described below. Standardized factor scores were used in all genetic analyses.

Analyses

GREML analyses utilized 796,125 autosomal markers that were obtained using the Illumina 1M platform on blood samples deposited at the Rutgers University Cell and DNA Repository. SNPs had to meet the following criteria for inclusion in the study: minor allele frequency >1%, genotyping call rate >99%, and a Hardy–Weinberg Equilibrium (HWE) p-value greater than 0.0001 among subjects of European descent (Palmer *et al.*, 2015b). Genomic principal component analysis was used to identify and retain subjects of European ancestry by including samples from the haplotype mapping project (HAPMAP; i.e., CEPH, Yoruban, Han Chinese, and Japanese) as ancestral reference groups in the genetic principal component analysis in SNP and Variation Suite (Palmer *et al.*, 2015a, SNP & Variation Suite [Version 8.3.4]).

GREML analyses were implemented in the GCTA software (Lee *et al.*, 2012, Yang *et al.*, 2010). Genomic relatedness matrices (GRM) were derived for all individuals and used to

describe all phenotypes. Analyses were limited to individuals who were no more related than second cousins (i.e., GRM-cutoff of 0.05 was imposed) so that estimates are not confounded with shared environmental effects and/or causal variants that are not tagged by the SNPs but captured by pedigree information. Reported estimates of variance captured by the SNPs (for DSM-IV symptoms) are transformed to that on the underlying scale using prevalence rates from cohort 2 of the Australian Twin study, which has levels of externalizing problems similar to general population samples in the US (Lessov *et al.*, 2004). Additive genetic effects (h^2_{SNP}) on individual FTND symptoms and the factor scores are un-scaled, and will be most relevant to a population with the same distribution of liability. All analyses controlled for age, gender, study of origin, and the first five genomic principal components derived from all subjects of European ancestry. For all GREML analyses, factor scores were standardized (Mean (M)=0, standard deviation (SD)=1)).

Results

Model-Building Phase - Factor Analysis of DSM-IV and FTND symptoms

EFA (N=1180 randomly selected smokers) suggested two latent variables (see Table 1 and Figure 1; eigenvalues (7.301, 1.596, 0.901, 0.644, 0.523, 0.441, 0.389, 0.358, 0.290, 0.254, 0.146, 0.131, and 0.026) that accounted for 56% of the variance across all items. The observed factors each accounted for at least 27% of the variance in a single item. Based on the RMSEA, CFI, TLI, and interpretability of factor loadings, the two-factor model provided the best fit for the data. The EFA results indicated that four DSM-IV symptoms (Tolerance, Withdrawal, ‘Activities foregone’, and ‘Continued use despite problems’) cross-loaded (i.e., 2+ symptoms loaded at 0.300 or higher on multiple factors) on the first and second factors. The first factor (referred to as the DSM-IV factor) was indicated by all seven DSM-IV criteria. Factor 2 (referred to as the FTND factor), was indicated by the FTND items and the four cross-loading DSM-IV items. Confirmatory factor analysis (N=1166 smokers) in the remainder of the sample indicated that the two-factor model with cross-loading items suggested by the EFA (i.e., CFA Model-I) provided a good fit to the data with an RMSEA value of 0.046 and CFI and TFI values above 0.95 (see Table 1). We also compared the fit of the two-factor model with cross-loading items to the fit of a model with no cross-loading items (i.e., CFA-Model-II). The overall model fit of the two factor model with no cross-loading items (CFA-Model-II) indicated that it provided a less parsimonious fit to the data when compared to the fit of CFA-Model-I, which included cross-loading items (i.e. the RMSEA increased (RMSEA = 0.015) and the CFI decreased [CFI = -0.018] for CFA-Model-II) (Cheung and Rensvold, 2002).

Data Extraction Phase – Factor Scores of Best-fitting Models

Based on the results from the model-building phase of our approach, factor scores were extracted using the full dataset. Factor loadings in the CFA (based on CFA-Model-I from the model-building component) using the entire dataset ranged from 0.33 to 0.93 indicated that the factors captured between 10%–80% of the variance in a given symptom (see standardized factor loadings in Figure 1). The observed correlation between the DSM-IV and FTND factors in the full sample was 0.771 [95% CI: 0.712, 0.831]. Given this high correlation, we fitted genetic models using a higher-order factor, “NIC DEP”. The higher-

order factor was obtained by restructuring Final-CFA-Model-I into Final-CFA-Model-I with higher-order factor, such that “NIC DEP” is indicated by the covariance between the FTND and DSM factors (see Figure 1). Note that the fit of Final-CFA-Model-I without the higher order factor is identical to the fit of Final-CFA-Model-I with the higher-order factor.

Variance in ND attributable to genotyped SNPs

Table 2 presents the univariate additive effect of genotyped SNPs on individual DSM-IV and FTND symptoms and the NIC DEP factor extracted from CFA-Final Model-I *with higher-order factor*. Additive genetic effects varied from 0–67%. The largest effect on DSM-IV items was 67% for nicotine withdrawal. The largest significant effect amongst FTND items was 32% for “Difficulty restraining from smoking in places where it is forbidden”. Focusing on the composite phenotype derived in the data extraction phase, our examination of additive genetic effects on the higher-order factor (“NIC DEP”), representing the shared variance across both factors, indicated a SNP-heritability of 33% (SE=0.142, $p=0.009$).

Post hoc analyses – Examination of DSM and FTND factors without cross-loading items

To further understand the nature of the overlap between the FTND and DSM factors, genetic analyses were also conducted on separate FTND and DSM factor scores obtained from a CFA on the full dataset using the structure of CFA-Model-II from the model-building component. This allowed us to examine genetic effects and genetic covariance between the factors without them having cross-loading items. This was achieved using univariate and bivariate GREML models fitted to DSM-IV and FTND factors derived from the full data (i.e., CFA-Final-Model-II [$\chi^2 = 680.48$, $df = 64$, $P < 0.0001$, RMSEA=0.064 [0.060, 0.068], CFI=0.953, TLI=0.943]). Presented in Table 2, additive genetic influences on the DSM-IV factor from CFA-Final-Model-II accounted for 29% (SE=0.141, $p=0.017$) of the variance of the factor. Likewise, additive genetic effects accounted for 25% (SE=0.142, $p=0.034$) of the variance of the FTND factor from CFA-Final-Model-II. The phenotypic correlation between the DSM-IV and FTND factors without cross-loading items was 0.752 [95% CI: 0.71–0.79]). The FTND and DSM-IV factors from CFA-Final-Model-II were highly correlated ($r_{g-SNP} = 1.00$ (SE=0.084, $p < 0.001$), suggesting shared additive genetic influences on DSM-IV and FTND symptoms of ND. These findings indicated shared genetic effects across indicators of ND when reflected separately by each measure.

Discussion

This is the first study of its kind to use GREML to parse the additive genetic variance associated with ND as a multidimensional phenotype. The study also indicates the extent to which additive genetic influences on two notable indices of nicotine dependence overlap. In doing so, we provide additional evidence to suggest that common genetic variance is the driving force behind the underlying liability to nicotine dependence (i.e., the NIC DEP factor) as jointly measured by the DSM-IV and the FTND. We used powerful genome-wide tools to estimate the SNP heritability across multiple indices of nicotine dependence in order to examine etiological and phenotypic heterogeneity. The additive effect of common SNPs for factor scores derived from our two gold-standard ND measures, the NIC DEP factor (33%), DSM-IV ND (29%) and FTND (25%), were each moderately high. In contrast, the

additive effect of common SNPs for more discrete aspects of ND varied widely, ranging from 0–67% with the highest estimates associated with DSM-IV withdrawal (67%), DSM-IV tolerance (29%), DSM-IV difficulty quitting smoking (26%), DSM-IV foregoing other activities in order to smoke (39%), FTND difficulty refraining from smoking when it is forbidden (32%), and FTND smoking while ill (29%). First and foremost, these data reveal the utility of common SNPs to index the genetic liability to ND and build on a limited but growing body of work supporting GREML as a useful tool in studies of substance dependence and other complex traits (Palmer *et al.*, 2015a, Vrieze *et al.*, 2013). In addition, these findings support the preponderance of evidence that ND is a multi-dimensional phenotype that is influenced by common genetic variants. Importantly, despite the range of the heritability estimates across DSM-IV and FTND measures, the effects attributed to common SNPs correlated highly across the DSM-IV and FTND derived factors. Thus, although phenotypic and etiological heterogeneity are present, our data demonstrate that discrete phenotypic aspects of ND share common underlying genetic risk.

Our estimates are consistent with one prior study using GREML to estimate SNP heritability for DSM-III nicotine dependence diagnosis, which reported an aggregate SNP heritability of 36% in a sample of both related and unrelated individuals (Vrieze *et al.*, 2013). Together, these two studies have leveraged GREML to account for substantially larger genetic effects than prior genome-wide and candidate studies of smoking and ND. The effects of common SNPs appear to account for a relatively large portion of genetic effects detected using traditional twin modeling approaches suggesting that candidate and whole genome genetic association (including GREML) approaches focused on common SNPs will have utility for contributing to the understanding of ND genetic risk pathways. As illustrated by the present article and recent addiction studies (Palmer *et al.*, 2015a, Palmer *et al.*, 2015b, Vrieze *et al.*, 2013), GREML holds unique promise for advancing the field to the extent that researchers find novel ways to parse the observed genetic variance to relevant genetic loci (either by function or prior knowledge). Likewise, the ability to examine the aggregate genetic effect allows for the explicit testing of the assumptions of GWAS. For example, Palmer *et al.* (2015b), recently confirmed the assumption of genetic homogeneity across DSM-IV indicators of alcohol dependence using genome-wide SNPs (Palmer *et al.*, 2015b), a question which future studies can help address for ND.

Our findings that SNP heritability estimates varied from very low to relatively high across various discrete ND measures are broadly consistent with prior work estimating heritability of item level ND measures using traditional twin approaches. Work from Lessov and colleagues (2004) examining genetic influences on DSM-IV criteria and two FTND items—time to first cigarette in the morning and CPD—found that additive genetic influences ranged from at least moderately (DSM-IV give up activities to smoke and DSM-IV smoking despite problems) to highly heritable (DSM-IV tolerance, DSM-IV withdrawal, FTND time to first cigarette, and FTND cigarettes per day) across the various indices. Although estimates differed across the two studies and overall tended to be higher in the Lessov biometric twin study, together findings support that genetic effects vary across ND dimensions and that withdrawal and tolerance may be the most genetically influenced indicators of ND in adults. Some of the genetic variance influencing these particular physiological ND traits is likely to be driven, at least partially, by the widely-replicated

genetic effects found within pharmacodynamic or pharmacokinetic genetic pathways, such as the relatively large effects found for nicotinic receptor gene variation on ND-related traits (e.g. Bloom *et al.*, 2014, Hancock *et al.*, 2015). However, our estimates suggest that genetic variation due to common SNPs accounts for a much larger portion of the variance in ND (25–33%) than has been accounted for in prior genetic association work (typically <5% with any particular gene or gene cluster). This discrepancy suggests that other genes and gene systems, beyond even those with the largest and most well-replicated effects, are also relevant to genetic models of ND.

Although prior work has supported shared genetic variance among ND phenotypes (Edwards *et al.*, 2011), importantly, ours is the first report to examine the correlation of genetic effects *due to common SNPs* across dimensions of ND. Although heritability estimates varied across ND measures, common SNP effects correlated highly across the DSM-IV and FTND factors suggesting that common genetic risk underlies these correlated but distinct ND phenotypes. An implication of this finding is that by employing multiple phenotypes that share a known underlying risk dimension, we would improve both detection and replication of genetic association with common SNPs. The current findings suggest that, despite measurement issues in ND research, individual differences with respect to the DSM-IV and FTND symptoms are accounted for by the same set of common SNPs. While the FTND and DSM-IV clearly demonstrate differences in sensitivity to the various domains that make up the broader ND phenotype (Koob and Volkow, 2010), our findings of evidence for overlapping genetic effects across the two measures are indicative exclusively of common genetic substrates that are collectively a set of ND risk factors. However, they do not specifically delineate biological mechanisms/pathways (e.g., neuronal tissues and circuits) that are involved in ND. That said, our data support the presence of an underlying construct for ND and indicate that FTND and DSM-IV criteria are alternate manifestations of the set of behavioral, psychiatric, and physical problems that can occur in relation to dependence on nicotine. As such, they share common risk factors and it is plausible that genetic associations from studies examining these various ND measures can be compared meaningfully.

Limitations

While the current study was able to maximize power by utilizing the largest, publicly available, and genetically-informed sample on drug addiction, there are several limitations. Power analyses based on the current sample size indicated that we were well powered (over 90%) to detect heritability estimates of 0.40 or greater, which is one-half of the heritability of ND observed in twin studies. Despite the large standard errors, the observed point estimates have been shown to be consistent across sample sizes (Stanton-Geddes *et al.*, 2013). Future studies with larger samples are needed to replicate these findings and to confirm the observed effects on individual symptoms of the DSM-IV and FTND scales. It should also be noted that the current findings are limited to common genetic variants; however, recent findings implicate a possible role for rare variants (Olfson *et al.*, 2015, Vrieze *et al.*, 2014), suggesting the need for future whole-genome studies incorporating common and rare polymorphisms. Readers should also consider that SAGE is a selected sample of individuals (i.e., drug addicted cases and controls). As such, these findings may not generalize to community and general population samples that might show different

patterns of drug use and dependence problems. In most analyses, we utilized population-based prevalence rates to adjust the observed genetic variance on the liability scale so that our estimates of SNP heritability are generalizable to an unselected population; despite these efforts, we conservatively expect that the SNP heritability may be lower in the general (non-ascertained) population (Lee *et al.*, 2012). Future research, using larger epidemiologic samples recruited for population representativeness rather than substance use is needed to replicate the present findings. Lastly, although the SAGE study employed gold standard ND measures (DSM-IV and FTND) in order to capture assessments that would be most generalizable across studies, discrete aspects of ND such as withdrawal severity, tolerance, and difficulties with quitting were assessed with a single self-report item. To the extent these are important to the genetic etiology of multidimensional ND, more detailed and comprehensive assessments of these ND facets should be tested in future studies of the effects of common SNPs on ND. Further, additional ND-related phenotypes, such as a biological assessment of smoke exposure, would be informative as studies have suggested that biological phenotypes like carbon monoxide levels may capture unique genetic variance to behavioral ND-related measures such as frequency of smoking (Bloom *et al.*, 2014).

Summary

In sum, our data suggest that between 26–33% of the variance in gold standard ND measures is attributable to common SNPs. Further although estimates of the additive effects of these common SNPs varied across the multidimensional ND phenotype, the genetic variance across multiple ND indicators was attributable to same set of common SNPs suggesting that discrete aspects of ND and subsets of dependent smokers share a common genetic etiology.

Acknowledgments

Financial support:

Funding support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment, and Health Initiative [GEI] (U01HG004422). SAGE is one of the genome-wide association studies funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by the Collaborative Study on the Genetics of Alcoholism (COGA; U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (COGEND; P01 CA089392), and the Family Study of Cocaine Dependence (FSCD; R01 DA013423). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), the National Institute on Alcohol Abuse and Alcoholism, the National Institute on Drug Abuse, and the NIH contract “High throughput genotyping for studying the genetic contributions to human disease” (HHSN268200782096C). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1 through dbGaP accession number phs000092.v1.p1. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs.

Funding for authors of this study was provided by DA033302 (Bidwell), DA023134 (Knopik), and AA021113 (Palmer). NIAAA and NIDA had no role in the study design, collection, analysis, or interpretation of the data, writing the manuscript, or the decision to submit the paper for publication.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-III-R.
3. American Psychiatric Association; Washington, DC: 1987. rev. ed

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-IV. 4. American Psychiatric Association; Washington, DC: 1994.
- Baker TB, Weiss RB, Bolt D, von Niederhausern A, Fiore MC, Dunn DM, Piper ME, Matsunami N, Smith SS, Coon H, McMahon WM, Scholand MB, Singh N, Hoidal JR, Kim S-Y, Leppert MF, Cannon DS. Human neuronal acetylcholine receptor A5-A3-B4 haplotypes are associated with multiple nicotine dependence phenotypes. *Nicotine & Tobacco Research*. 2009
- Bloom AJ, Hartz SM, Baker TB, Chen L, Piper ME, Fox L, Martinez M, Hatsukami D, Johnson EO, Laurie CC, Saccone NL, Goate A, Bierut LJ. Beyond Cigarettes Per Day. A Genome-Wide Association Study of the Biomarker Carbon Monoxide. *Annals of the American Thoracic Society*. 2014; 11:1003–1010. [PubMed: 25072098]
- Breslau N, Johnson EO. Predicting smoking cessation and major depression in nicotine-dependent smokers. *American Journal of Public Health*. 2000; 90:1122–1127. [PubMed: 10897192]
- Broms U, Silventoinen K, Madden PAF, Heath AC, Kaprio J. Genetic Architecture of Smoking Behavior: A Study of Finnish Adult Twins. *Twin Research and Human Genetics*. 2006; 9:64–72. [PubMed: 16611469]
- Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI, Reich T, Schmidt I, Schuckit MA. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *Journal of Studies on Alcohol*. 1994; 55:149–158. [PubMed: 8189735]
- Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*. 2002; 9:233–255.
- Dijkstra A, Tromp D. Is the FTND a measure of physical as well as psychological tobacco dependence? *Journal of substance abuse treatment*. 2002; 23:367–74. [PubMed: 12495799]
- Edwards AC, Maes HH, Pedersen NL, Kendler KS. A population-based twin study of the genetic and environmental relationship of major depression, regular tobacco use and nicotine dependence. *Psychological Medicine*. 2011; 41:395–405. [PubMed: 20406522]
- Hancock DB, Reginsson GW, Gaddis NC, Chen X, Saccone NL, Lutz SM, Qaiser B, Sherva R, Steinberg S, Zink F, Stacey SN, Glasheen C, Chen J, Gu F, Frederiksen BN, Loukola A, Gudbjartsson DF, Bruske I, Landi MT, Bickeboller H, Madden P, Farrer L, Kaprio J, Kranzler HR, Gelernter J, Baker TB, Kraft P, Amos CI, Caporaso NE, Hokanson JE, Bierut LJ, Thorgeirsson TE, Johnson EO, Stefansson K. Genome-wide meta-analysis reveals common splice site acceptor variant in *CHRNA4* associated with nicotine dependence. *Translational Psychiatry*. 2015; 5:e651. [PubMed: 26440539]
- Heatherton TF, Kozlowski LT, Frecker RC, Rickert W, Robinson J. Measuring the Heaviness of Smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. *British Journal of Addiction*. 1989; 84:791–800. [PubMed: 2758152]
- Hettema JM, Corey LA, Kendler KS. A multivariate genetic analysis of the use of tobacco, alcohol, and caffeine in a population based sample of male and female twins. *Drug and Alcohol Dependence*. 1999; 57:69–78. [PubMed: 10617315]
- Hudmon KS, Marks JL, Pomerleau CS, Bolt DM, Brigham J, Swan GE. A multidimensional model for characterizing tobacco dependence. *Nicotine & Tobacco Research*. 2003; 5:655–664. [PubMed: 14577982]
- Hughes, JR. Those who continue to smoke: Is achieving abstinence harder and do we need to change our interventions?. Bethesda, MD: 2003. The case for hardening the target. *Smoking and Tobacco Control Monograph No. 15*. (NIH Publication No. 03–5370)
- Hughes JR, Oliveto AH, Riggs R, Kenny M, Liguori A, Pillitteri JL, MacLaughlin MA. Concordance of different measures of nicotine dependence: Two pilot studies. *Addictive Behaviors*. 2004; 29:1527–1539. [PubMed: 15451122]
- Kandel D, Schaffran C, Griesler P, Samuolis J, Davies M, Galanti R. On the Measurement of Nicotine Dependence in Adolescence: Comparisons of the mFTQ and a DSM-IV–Based Scale. *Journal of Pediatric Psychology*. 2005; 30:319–332. [PubMed: 15863429]
- Kaprio JA, Koskenvuo M, Sarna S. Cigarette smoking, use of alcohol, and leisure-time physical activity among same-sexed adult male twins. *Progress in Clinical and Biological Research*. 1981; 69:37–46. [PubMed: 7198250]

- Koob GF, Volkow ND. Neurocircuitry of addiction. *Neuropsychopharmacology* : official publication of the American College of Neuropsychopharmacology. 2010; 35:217–38. [PubMed: 19710631]
- Koopmans J, Slutske W, Heath A, Neale M, Boomsma D. The Genetics of Smoking Initiation and Quantity Smoked in Dutch Adolescent and Young Adult Twins. *Behavior Genetics*. 1999; 29:383–393. [PubMed: 10857244]
- Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*. 2012; 28:2540–2542. [PubMed: 22843982]
- Lessov CN, Martin NG, Statham DJ, Todorov AA, Slutske WS, Bucholz KK, Heath AC, Madden PAF. Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychological Medicine*. 2004; 34:865–879. [PubMed: 15500307]
- Loukola A, Wedenoja J, Keskitalo-Vuokko K, Broms U, Korhonen T, Ripatti S, Sarin A-P, Pitkaniemi J, He L, Häppölä A, Heikkilä K, Chou Y-L, Pergadia ML, Heath AC, Montgomery GW, Martin NG, Madden PAF, Kaprio J. Genome-wide association study on detailed profiles of smoking behavior and nicotine dependence in a twin sample. *Molecular psychiatry*. 2014; 19:615–624. [PubMed: 23752247]
- MacCallum, RC.; Browne, MW.; Sugawara, HM. Power analysis and determination of sample size for covariance structure modeling. American Psychological Association; US: 1996. p. 130-149.
- McGue M, Elkins I, Iacono WG. Genetic and environmental influences on adolescent substance use and abuse. *American Journal of Medical Genetics*. 2000; 96:671–677. [PubMed: 11054776]
- Moolchan ET, Radzins A, Epstein DH, Uhl G, Gorelick DA, Cadet JL, Henningfield JE. The Fagerstrom Test for Nicotine Dependence and the Diagnostic Interview Schedule: Do they diagnose the same smokers? *Addictive Behaviors*. 2002; 27:101–113. [PubMed: 11800217]
- National Cancer Institute. Tobacco Control Monograph No. 20. U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute; Bethesda, MD: 2009. Phenotypes and Endophenotypes: Foundations for Genetic Studies of Nicotine Use and Dependence. NIH Publication No. 09-6366
- National Center for Biotechnology Information [NCBI]. [Accessed January 28, 2016] Study of Addiction: Genetics and Environment (SAGE). n.d. (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1)
- Olfson E, Saccone NL, Johnson EO, Chen LS, Culverhouse R, Doheny K, Foltz SM, Fox L, Gogarten SM, Hartz S, Hetrick K, Laurie CC, Marosy B, Amin N, Arnett D, Barr RG, Bartz TM, Bertelsen S, Borecki IB, Brown MR, Chasman DI, van Duijn CM, Feitosa MF, Fox ER, Franceschini N, Franco OH, Grove ML, Guo X, Hofman A, Kardina SLR, Morrison AC, Musani SK, Psaty BM, Rao DC, Reiner AP, Rice K, Ridker PM, Rose LM, Schick UM, Schwander K, Uitterlinden AG, Vojinovic D, Wang JC, Ware EB, Wilson G, Yao J, Zhao W, Breslau N, Hatsukami D, Stitzel JA, Rice J, Goate A, Bierut LJ. Rare, low frequency and common coding variants in CHRNA5 and their contribution to nicotine dependence in European and African Americans. *Molecular psychiatry*. 2015
- Palmer RHC, Brick L, Nugent NR, Bidwell LC, McGeary JE, Knopik VS, Keller MC. Examining the role of common genetic variants on alcohol, tobacco, cannabis and illicit drug dependence: genetics of vulnerability to drug dependence. *Addiction*. 2015a; 110:530–537. [PubMed: 25424661]
- Palmer RHC, McGeary JE, Heath AC, Keller MC, Brick LA, Knopik VS. Shared additive genetic influences on DSM-IV criteria for alcohol dependence in subjects of European ancestry. *Addiction*. 2015b; 110:1922–1931. [PubMed: 26211938]
- Payne TJ, Smith PO, McCracken LM, McSherry WC, Antony MM. Assessing nicotine dependence: a comparison of the Fagerstrom Tolerance Questionnaire (FTQ) with the Fagerstrom Test for Nicotine Dependence (FTND) in a clinical sample. *Addictive Behaviors*. 1994; 19:307–17. [PubMed: 7942248]
- Pergadia M, Heath A, Agrawal A, Bucholz K, Martin N, Madden PF. The Implications of Simultaneous Smoking Initiation for Inferences about the Genetics of Smoking Behavior from Twin Data. *Behavior Genetics*. 2006a; 36:567–576. [PubMed: 16477519]
- Pergadia ML, Heath AC, Martin NG, Madden PAF. Genetic analyses of DSM-IV nicotine withdrawal in adult twins. *Psychological Medicine*. 2006b; 36:963–972. [PubMed: 16749946]

- Piper ME, McCarthy DE, Baker TB. Assessing Tobacco Dependence: A Guide to Measure Evaluation and Selection. *Nicotine & Tobacco Research*. 2006; 8:339–351. [PubMed: 16801292]
- Piper ME, Piasecki TM, Federman EB, Bolt DM, Smith SS, Fiore MC, Baker TB. A Multiple Motives Approach to Tobacco Dependence: The Wisconsin Inventory of Smoking Dependence Motives (WISDM-68). *Journal of Consulting and Clinical Psychology*. 2004; 72:139–154. [PubMed: 15065950]
- Pomerleau OF, Collins AC, Shiffman S, Pomerleau CS. Why some people smoke and others do not: New perspectives. *Journal of Consulting and Clinical Psychology*. 1993; 61:723–731. [PubMed: 8245270]
- Shiffman S, Waters AJ, Hickcox M. The Nicotine Dependence Syndrome Scale: A multidimensional measure of nicotine dependence. *Nicotine & Tobacco Research*. 2004; 6:327–348. [PubMed: 15203807]
- Bozeman, MT: Golden Helix, Inc; SNP & Variation Suite [Version 8.3.4] SNP & Variation Suite (Version 8.3.4) [Software]. Available from <http://www.goldenhelix.com>
- Stanton-Geddes J, Yoder JB, Brikin R, Young ND, Tiffin P. Estimating heritability using genomic data. *Methods in Ecology and Evolution*. 2013; 4:1151–1158.
- Swan GE, Carmelli D, Cardon LR. Heavy consumption of cigarettes, alcohol and coffee in male twins. *Journal of Studies on Alcohol*. 1997; 58:182–190. [PubMed: 9065896]
- Swan GE, Carmelli D, Rosenman RH, Fabsitz RR, Christian JC. Smoking and alcohol consumption in adult male twins: Genetic heritability and shared environmental influences. *Journal of Substance Abuse*. 1990; 2:39–50. [PubMed: 2136102]
- True WR, Xian H, Scherrer JF, Madden PAF, Bucholz KK, Heath AC, Eisen SA, Lyons MJ, Goldberg J, Tsuang M. Common genetic vulnerability for nicotine and alcohol dependence in men. *Archives of General Psychiatry*. 1999; 56:655–661. [PubMed: 10401514]
- Vrieze SI, Feng S, Miller MB, Hicks BM, Pankratz N, Abecasis GR, Iacono WG, McGue M. Rare nonsynonymous exonic variants in addiction and behavioral disinhibition. *Biological psychiatry*. 2014; 75:783–9. [PubMed: 24094508]
- Vrieze SI, McGue M, Miller MB, Hicks BM, Iacono WG. Three Mutually Informative Ways to Understand the Genetic Relationships Among Behavioral Disinhibition, Alcohol Use, Drug Use, Nicotine Use/Dependence, and Their Co-occurrence: Twin Biometry, GCTA, and Genome-Wide Scoring. *Behavior Genetics*. 2013; 43:97–107. [PubMed: 23362009]
- Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*. 2010; 42:565–569. [PubMed: 20562875]
- Yang, J.; Lee, S.; Goddard, M.; Visscher, P. Genome-Wide Complex Trait Analysis (GCTA): Methods, Data Analyses, and Interpretations. In: Gondro, C.; van der Werf, J.; Hayes, B., editors. *Genome-Wide Association Studies and Genomic Prediction*. Humana Press; 2013. p. 215-236.

Figure 1.

Confirmatory Factor Analysis of DSM-IV and FTND symptoms in the entire SAGE sample (Final-CFA-Model-I with Higher-order factor). Each standardized path loading (with 95% confidence interval) could not be dropped from the model without a significant loss in model fit. *Abbreviations:* DSM1-Tolerance; DSM2-Withdrawal; DSM3-Using more than intended; DSM4-Unsuccessful attempts to cut down; DSM5-Great time spent using/recovering; DSM6-Activities foregone; DSM7-Continued use despite problems; FTND1-Time to first cigarette; FTND2- Difficulty refraining; FTND3- Most difficult cigarette to quit; FTND4- Cigarettes per day; FTND5- Timing of greatest smoking frequency; FTND6- Smoking while ill. Notation: * - The unstandardized path loading from DSM to NIC DEP was fixed to 1.00 in the Mplus model, thus the standardized loading presented does not have standard error. The product of the standardized loadings on NIC DEP is equal to the phenotypic correlation between the two factors ($r_{\text{DSM-FTND}} = 0.545$ [95% CI: 0.50–0.60]).

Table 1

Model fit and loadings for 2-factor EFA/CFA

	Model Building			Data Extraction	
	EFA (N=1180)			Final CFA (N=2346)	
	1-factor	2-factor		Model-I	Model-II with higher-order factor
Chi-Square Test	654.484	163.689		337.23	372.965
DF	65	53		64	78
P-Value	<0.0001	<0.0001		<0.0001	<0.0001
RMSEA [90% CI]	0.088 [0.082, 0.094]	0.042 [0.035, 0.049]		0.061 [0.054, 0.067]	0.047 [0.043, 0.052]
CFI	0.907	0.982		0.96	0.976
TLI	0.888	0.974		0.951	0.969

Table describes the fit indices and loadings from the exploratory (EFA) and confirmatory factor analyses (CFA) conducted using random, independent subsets of the full dataset, and the extraction using the full dataset. Two confirmatory factor models were fitted (Model I - based on the 2-factor EFA solution, Model II - a 2-factor model that had separate factors each indicated by DSM/FTND items only). Abbreviations: CFI - Comparative Fit Index, DF - Degrees of freedom, N/A - Not applicable, RMSEA - Root Mean Square Error of Approximation, TLI - Tucker-Lewis index.

Table 2

Proportion of variance explained by genome-wide SNPs across ND measures

Phenotype	Current Paper			Prior twin studies
	h^2_{SNP1}	s.e.	p-value	h^2^a
<i>DSM-IV</i>				
Sx 1: Tolerance	0.292	0.186	0.057	0.73
Sx 2: Withdrawal	0.673	0.231	0.002	0.53
Sx 3: Using more than intended	0.135	0.207	0.253	0.62
Sx 4: Unsuccessful attempts to cut down	0.265	0.24	0.131	0.57 ^b
Sx 5: Great time spent using/recovering	<0.001	0.744	0.5	0.45
Sx 6: Activities foregone	0.392	0.472	0.208	0.26
Sx 7: Continued use despite problems	<0.001	0.243	0.5	0.39
<i>FTND</i>				
Sx 1: Time to first cigarette	0.154	0.167	0.172	0.68
Sx 2: Difficulty refraining	0.323	0.167	0.023	N/A
Sx 3: Most difficult cigarette to quit	<0.001	0.177	0.496	N/A
Sx 4: Cigarettes per day	0.078	0.166	0.318	0.7
Sx 5: Timing of greatest smoking frequency	0.17	0.165	0.142	N/A
Sx 6: Smoking while ill	0.29	0.169	0.039	N/A
<i>Composite Measures from Data Extraction Phase</i>				
Higher-order factor (NIC DEP) ^c	0.332	0.142	0.009	N/A
FTND Factor ^d	0.291	0.141	0.017	
DSM-IV Factor ^d	0.254	0.142	0.034	N/A

Table showing the univariate SNP-heritability of DSM-IV and FTND indicators while scaling the estimates to population levels observed in Lessov et al. 2004. Note that effects on composite measures have not been adjusted for variance on the underlying scale. Notation:

^a Univariate twin heritability estimates from Lessov et al., 2004;

^b estimates averaged across genders.

^c factor scores extracted from Final-CFA-Model-I with higher-order factor.

^d factor scores extracted from CFA-Final-Model-II.

Abbreviations: N/A – Not available.